DOCUMENT RESUME

ED 190 616 TM 800 410

AUTHOR Perrin, David W.; Kerasotes, Dean L.
TITLE The Effect of Asterisks as an Attention Focusing Device on Test Performance of Undergraduate Students.
PUB DATE [79]
NOTE 12p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Difficulty Level; Higher Education; *Objective Tests; *Response Style (Tests); Scores; Test Items; *Test Wiseness; Undergraduate Students; Visual Stimuli

ABSTRACT

It was hypothesized that using asterisks as attention focusing devices would cause students to read all asteriked test items more carefully and would improve test scores of undergraduate education students. Sixty-three undergraduates majoring in elementary or special education were administered a 36-item objective test. Asterisks were used to focus students' attention on randomly selected items: one third of the tests had no asterisks, one third had asterisks with 25% of the items, and one third had asterisks with 50% of the items. The results of the analysis of variance are reported in tabular form. No differences were observed among the three forms of the test. Findings revealed that the presentation of asterisks with items of all difficulties failed to function effectively as an attention focusing item, and thereby increase test scores. (Author/GSK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
D. PERRIN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Effect of Asterisks as an Attention Focusing Device on Test Performance of Undergraduate Students

David W. Perrin and Dean L. Kerasotes
Oklahoma State University

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED190616

Abstract

A response-set may cause students to respond incorrectly to difficult test items which appear easy on initial reading. Presentation of item difficulties has been demonstrated effective in negating this postulated response-set. The primary hypothesis was that randomly assigning asterisks as an attention focusing device would cause students to read all asterisked items (regardless of difficulty) more carefully and, thereby, improve their test scores. No differences were observed among three forms (0%, 25%, 50% asterisks) on a 36 item test containing 12 multiple-choice, true-false, and completion items. Results were interpreted to support possible operation of the response-set only with certain more difficult items.

TM 800 410

The Effect of Asterisks as an Attention Focusing Device on Test Performance of Undergraduate Students

David W. Perrin and Dean L. Kerasotes
Oklahoma State University

Students may often respond hastily and incorrectly to difficult objective test items which appear easy on an initial reading. It has recently been suggested that, given item difficulties, students can be encouraged to read these items more carefully and, thereby, increase their test scores (Huck, 1978). The purpose of this research was to determine if some means other than presentation of item difficulties might be used to encourage students to read certain items more carefully.

Hasty and incorrect responses may be considered a response-set which has been defined as "a habit or a momentary set causing the subject to earn a different score from the one he would earn if the same items were presented in a different form" (Cronbach, 1970, p. 148). The existence of this response-set for graduate students in education has been substantiated by Huck (1978). Using three types of objective items (multiple choice, true-false, completion) and two item forms (items presented with and without difficulty information), Huck (1978) concluded that students administered items accompanied with difficulty estimates obtained significantly higher scores than students administered items without difficulty estimates. This conclusion strongly implies the existence of a careless reading response-set for at least some students on certain objective test items.

The hypothesis of this study was that the use of asterisks as an attention focusing device would improve test scores of undergraduate students in education. Additionally, the authors were curious if the postulated response-set

operated for items of all difficulties (not just "hard" items). Therefore, asterisks were randomly assigned to items in an attempt to demonstrate that a simple attention focusing device could encourage students to read asterisked items more carefully and, thereby, improve their test scores. Such an improvement would further substantiate the existence of the postulated response-set.

METHOD

Subjects

The subjects used in this study were 63 undergraduate students enrolled in a child development course at a major state university. The purpose of the course was to introduce prospective teachers to the cognitive, affective, and physical development of children. The students enrolled in the course were primarily majoring in either elementary education or special education.

Test

The 36 item test consisted of 12 multiple choice (MC), 12 true-false (TF), and 12 completion or fill-in-the-blank (FITB) items. Asterisks were used to focus students' attention on randomly selected items. One-third of the tests contained no asterisks, one-third presented asterisks with 25% of the items, and one-third presented asterisks with 50% of the items. Items accompanied by an asterisk were randomly chosen within item type. All items with an asterisk in the 25% forms were also accompanied by an asterisk in the 50% forms. Within each attention level (0%, 25%, and 50% asterisks), three sequences of item types (MC, TF, FITB; TF, FITB, MC; FITB, MC, TF) were employed. Thus, nine forms of the test were prepared by combining the three attention levels with the three sequences of item type. All items for the test, which was the first examination of the semester, were chosen or written by the instructor specifically for the course.

Procedure

The nine forms of the test were arranged in random order and distributed to 64 students during the regularly scheduled class period.[1] Two sets of instructions, one for the test forms without asterisks and one for the test forms containing asterisks, were placed on the cover-sheets for the tests. On test forms containing asterisks, the following instructions were used:

> This test contains multiple-choice, true-false, and completion items. Your score will be the number of items answered correctly (there will be no correction for guessing) so you should answer every item. Items denoted with asterisks have been difficult for students in the past.

Instructions on test forms not containing asterisks were identical except the underlined sentence was omitted. After the tests were distributed, the instructor announced that each student should carefully read the instructions prior to beginning. The instructor reported no evidence that indicated the students were aware that different test forms had been used.

Design

The experimental design employed was analogous to that used by Huck (1978). This design is often identified as a Lindquist (1953) Type IV design and is termed a repeated measures Latin square design, Plan 9 by Winer (1971). Order and item type were considered within-subjects effects. Attention level and sequence were considered between-subjects effects. The response measures were the scores of the students for each item type.

RESULTS AND CONCLUSIONS

The results of the analysis of variance are presented in Table 1. The F statistic used to test the effect of item type was statistically significant ($p < .001$). No other statistically significant values of F were observed.

---

[1] The test paper for one student was randomly deleted from the group which received the test form with 25% asterisks and item sequence TF, FITB, MC. Scores for students absent the day of the test were not included in the study.

Thus, it was concluded that, except for item type, none of the experimental variables influenced student test scores.

- -

Insert Table 1 about here

- -

Means, median difficulty (p) levels, standard deviations, and Kuder-Richardson 20 reliability estimates for the total test at each attention level are presented in Table 2. These statistics support the conclusion that attention level, the variable of primary interest, did not influence test scores. The mean scores and median difficulty (p) levels for each item type are also presented in Table 2. A Tukey (a) test among the main effect means for item type (MC = 6.59, TF = 8.60, FITB = 5.38) indicated that all pairwise comparisons were different ($p < .05$).

- -

Insert Table 2 about here

- -

DISCUSSION

Unlike the presentation of item difficulties as an attention focusing device, the presentation of asterisks with items failed to increase student test scores. Several possible explanations for this apparent contradiction have occurred to the authors. It may be that undergraduate students in a child development course read the test questions on the first exam more carefully than graduate students in a statistics course did on the final exam. Or it could be that the reuse of test items in the Huck (1978) study actually induced carelessness (negative response-set) among the graduate students not presented with item difficulties. Such a response-set would be expected to lower the scores of those students not receiving item difficulties. Huck (1978)

does not directly address this possibility and the reported nonsignificant difference ($z = .91$, $p = .36$) between the mean item difficulty level for non-random groups of previous students ($\bar{p} = .74$) and current students not presented item difficulties ($\bar{p} = .67$) does not eliminate this possibility. However, neither of these possible explanations seem highly likely.

A more probable explanation may be made if one assumes that the postulated response-set operates only for certain difficult items. Huck reported that the group presented with item difficulties (p group) performed better than the group not presented with item difficulties (non-p group) on 12 of the 15 most difficult items on his exam. These 15 items (5 MC, 5 TF, and 5 FITB) represented the most difficult 20% of each type of item on his exam. Let us assume that the three items which were more difficult for the p group cancelled the effect of three of the 12 items which were less difficult for that group and that the difference between mean test scores (approximately 5.0 points) was caused by the postulated response-set operating within the non-p group on the 9 remaining difficult items. Then, the mean difficulty value for these 9 items would have been 5/9 (or approximately .55) higher in the p group than in the non-p group. Thus, the approximate difference between the p group and the non-p group in the Huck study may be expressed as

$$D \doteq k(P)(r)(d)$$

where D = mean score difference between p and non-p groups,

k = number of items on the test,

P = proportion of hard items (those for which difficulty indices were presented),

r = proportion of hard items for which the postulated response-set operated, and

d = difference in mean difficulty between the p and non-p groups on those items for which the postulated response-set operated.

The actual difference (4.74) in the Huck study is closely approximated by

$$5 \doteq 75 \times 15/75 \times 9/15 \times 5/9 = 75 \times .20 \times .60 \times .55.$$

If a similar effect could be assumed for the presentation of asterisks, then the present authors might have expected

$$2.4 \doteq 36 \times .20 \times .60 \times .55$$

difference if the asterisks had been presented with the most difficult 20% of the items and the postulated response-set had operated for the same proportion of these items. Assignment of asterisks to the more difficult items and a difference in mean test scores of the size predicted above would have been more consistent with the logic and results of Huck's study.

A post-hoc analysis of the most difficult items from the present test was conducted. Ten items; three MC, three TF, and four FITB (two of these items were equally difficult) were included in this analysis. Paired comparisons of the difficulties of these ten items administered without asterisks were made to the difficulties of the same ten items administered under the 25% and 50% asterisks conditions. Of these 20 comparisons, eight indicated the item was easier when asterisked and ten indicated the item was more difficult when asterisked (two comparisons were tied). A binomial test suggested there was no difference between the asterisk and non-asterisk conditions even for the more difficult items.

In the authors' opinion, the results of this study do not eliminate the possibility of a response-set which causes students to incorrectly respond to difficult items which appear easy upon an initial reading. However, that response-set may be limited to just those items and may not operate with all difficult items. The present study suggests that the presentation of asterisks with items of all difficulties does not effectively function as an attention focusing device and, thereby, increase student test scores. This was

disappointing to the authors since the asterisk could be used in situations where item difficulties are not known (e.g., newly written items for which the instructor believed the postulated response-set would operate) or with younger students who did not understand item difficulties.

REFERENCES

Cronbach, L. J. Essentials of psychological testing (3rd ed.). New York: Harper & Row, 1970.

Huck, S. W. Test performance under the condition of known item difficulty. Journal of Educational Measurement, 1978, 15, 53-58.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

Table 1

Summary Table for Lindquist

Type IV Analysis of Variance

Source of Variation	df	SS	MS	F
Between Subjects	62	627.22		
Attention Level (AL)	2	1.03	.51	.05
Sequence (S)	2	2.87	1.43	.36
AL x S	4	21.70	5.43	.49
Error (b)	54	601.62	11.14	
Within Subjects	126	701.33		
Order (O)	2	.96	.48	.16
Item Type (IT)	2	330.42	165.21	53.13*
O x AL	4	3.51	.88	.28
IT x AL	4	5.58	1.39	.45
(O x IT)	2	9.28	4.64	1.49
(O x IT x AL)	4	15.70	3.93	1.26
Error (w)	108	335.81	3.11	
Total	188	1302.95		

*$p < .001$

Table 2

Item and Test Statistics by

Attention Level and Item Type

Attention Level	Mean	Median p-level	Standard Deviation	KR20
0% Level	20.76	57	5.49	.78
MC	6.33	50		
TF	8.90	69		
FITB	5.52	45		
25% Level	20.62	55	6.14	.83
MC	6.62	52		
TF	8.71	76		
FITB	5.29	46		
50% Level	20.24	53	5.11	.71
MC	6.48	51		
TF	8.24	76		
FITB	5.52	48		
Total	20.54	52	5.51	.81
MC	6.59	52		
TF	8.60	78		
FITB	5.38	48		